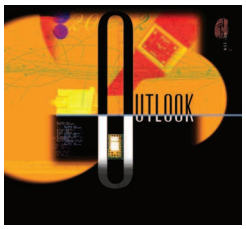
2001 Technology Roadmap for Semiconductors



The International Technology Roadmap for Semiconductors (ITRS) is a collaborative effort within the semiconductor industry to confront the challenges implicit in Moore's law. Representatives of the International Technology Working Groups for Design and Test outline some of the contributions made by 839 international experts as they sought to reach an industry-wide consensus on its R&D needs out to a 15-year horizon.

Alan Allan
Intel

Don Edenfeld
Intel

William H. Joyner Jr.
Semiconductor Research Corp.

Andrew B. Kahng
University of California, San Diego

Mike Rodgers
Intel

Yervant Zorian
LogicVision

For the past 40 years, the semiconductor industry has distinguished itself by the rapid pace of improvement in its products. This growth has resulted principally from the industry's ability to decrease exponentially the minimum feature sizes it uses to fabricate integrated circuits, commonly referred to as Moore's law. The most significant trend for society is the decreasing cost per function, which has led to significant improvements in productivity and quality of life through proliferation of computers, electronic communication, and consumer electronics.

Over the past two decades, the phenomenal increase in research and development investments has motivated industry collaboration and spawned many partnerships, consortia, and other cooperative ventures. The International Technology Roadmap for Semiconductors (ITRS) is an especially successful worldwide cooperation that presents an industry-wide consensus on the "best current estimate" of its R&D needs out to a 15-year horizon. As such, the Roadmap provides a guide to the efforts of companies, research organizations, and governments to improve the quality of R&D investment decisions made at all levels, and it has helped channel efforts to areas that truly need research breakthroughs.

Since its inception in 1992 as the *National Technology Roadmap for Semiconductors* (NTRS), the Roadmap's basic premise has been that scaling of microelectronics would continue to reduce the cost per function by 25 percent and promote market growth for integrated circuits by 15 percent annually. Thus, the Roadmap is put together in the spirit of a challenge: What technical capabilities does the industry need to develop to continue to follow Moore's law?

The semiconductor industry is increasingly sharing its research efforts via mechanisms such as consortia and collaborations with suppliers in a precompetitive environment. The ITRS identifies the principal technology needs to guide this shared research. It does this in two ways: by showing the targets that technology solutions currently under development need to meet and by indicating where there are no "known solutions" (of reasonable confidence) to continued scaling in some aspects of semiconductor technology. Because they clearly warn where historical progress trends might end if the industry doesn't achieve some real breakthroughs in the future, these latter indicators highlight serious and exciting challenges.

As the "Overall Roadmap Process and Structure" sidebar indicates, the 2001 Roadmap is notable because it was developed with truly international representation. The contributions outlined here represent but a small portion of this immense undertaking over the past two years.

Excerpts from the Executive Summary and three chapters of *The International Technology Roadmap for Semiconductors, 2001 edition*, International Sematech, Austin, Texas, 2001.

0018-9162/02/$17.00 © 2002 IEEE

CURRENT TRENDS

Historically, developers have recognized dynamic random access memory (DRAM) products as the technology drivers for the entire semiconductor industry. Prior to the early 1990s, logic technology, as exemplified by microprocessing units (MPUs), developed at a slower pace than DRAM technology. During the past few years, the development rate of new technologies used to manufacture microprocessors has accelerated, and DRAM product generation every three years at four times the previous density has become obsolete as a way to define technology nodes. (A technology node represents the creation of significant technology progress governed by the smallest feature printed—approximately 70 percent of the preceding node.)

Scaling

Microprocessor products are closing the historical half-pitch technology gap versus DRAM and are now driving the leading-edge lithography tools and processes—particularly with respect to the "printed in resist" and physical gate length. As Figure 1 shows, the 2001 Roadmap explicitly acknowledges that DRAM and microprocessor products share the technology leadership role, with MPU half-pitch closely tracking DRAM. In fact, MPU half-pitch will catch up to DRAM half-pitch in 2004; the previous edition of the ITRS had projected this convergence for 2015.

Despite the continuous reduction in feature size of about 30 percent every three years, the size of first DRAM product demonstration has continued to double every six years, an increase of about 12 percent per year. This increase in chip area has been necessary to accommodate 59 percent more bits/capacitors/transistors per year in accordance with Moore's law, historically doubling functions per chip every 1.5 to 2 years. However, to maintain the historical trend of reducing the cost/function ratio by 25 to 30 percent per year, the semiconductor industry must continuously enhance equipment productivity, increase manufacturing yields, use the largest wafer size available, and, most important, increase the number of chips available on a wafer.

Both the DRAM and MPU models depend upon achieving aggressive design and process improvement targets. If those targets slip, pressure will increase to print chip sizes larger than the present Roadmap predicts, or further slow the rate of Moore's law on-chip functionality. Either of these consequences will have a negative impact upon cost-per-function reduction rates—the classical measure of our industry's productivity, improvement, and competitiveness.

Overall Roadmap Process and Structure

The 2001 International Technology Roadmap for Semiconductors represents an attempt to incorporate broad international input to build the widest possible consensus on the semiconductor industry's future technology needs.

A corresponding International Technology Working Group (ITWG) writes each ITRS technology area chapter. The eight focus ITWGs correspond to typical subactivities that sequentially span product flow: design; test; process integration, devices, and structures; front-end processes; lithography; interconnects; factory integration; and assembly and packaging. Four crosscut ITWGs represent supporting activities that tend to individually overlap with product flow at multiple critical points: environment, safety, and health; defect reduction; metrology; modeling and simulation.

Each ITWG receives input from the Technology Working Groups (TWGs) in five geographical regions: Europe, Japan, Korea, Taiwan, and the US. One to two delegates represent each regional TWG on the corresponding ITWG. The regional TWGs are composed of experts from industry, including chip-makers as well as equipment and materials suppliers, government research organizations, and universities. In 2001, a total of 839 experts from the five regions volunteered their services in the 12 ITWGs. In addition, each TWG incorporates feedback gathered from an even larger community through sub-TWG meetings and public Roadmap workshops.

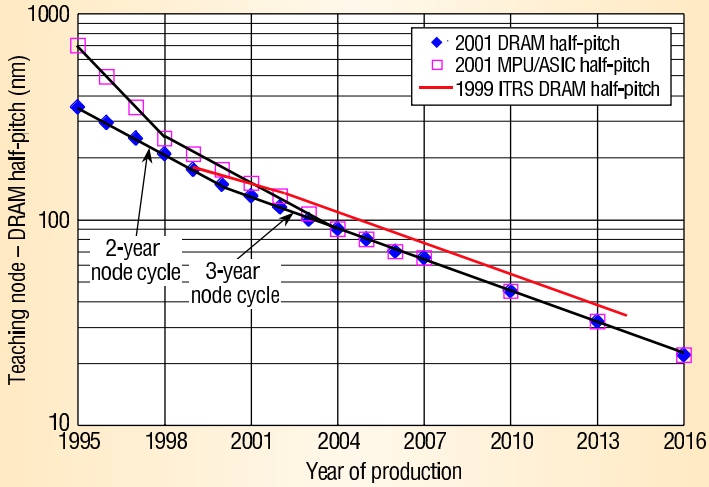


Figure 1. ITRS Roadmap acceleration continues with MPU half-pitch trends closely tracking DRAM and catching up to it in 2004 instead of 2015 as previously projected in the 1999 Roadmap.

Clock frequency

In addition to the need to increase functionality while exponentially decreasing cost per function, there is also a market demand for higher-performance, cost-effective products. Just as Moore's law predicts that functions per chip will double every 1.5 to 2 years to keep up with consumer demand, there is a corresponding demand for processing electrical signals at progressively higher rates. In the case of MPUs, processor instructions per second have also historically doubled every 1.5 to 2 years.

For MPU products, increased processing power, measured in millions of instructions per second (MIPs), is accomplished through a combination of raw technology performance (clock frequency) multiplied by architectural performance (instruc-

Table 1. Roadmap trends 2001-2016 for scaling, cost, power.

Production year	2001	2002	2003	2004	2005	2006	2007	2010	2013	2016
Chip performance characteristics										
DRAM half-pitch	130	115	100	90	80	70	65	45	32	22
MPU/ASIC half-pitch	150	130	107	90	80	70	65	45	32	22
MPU printed gate length	90	75	65	53	45	40	35	25	18	13
MPU physical gate length	65	53	45	37	32	28	25	18	13	9
On-chip local clock (MHz)	1,684	2,317	3,088	3,990	5,173	5,631	6,739	11,511	19,348	28,751
Maximum number wiring levels	7	8	8	8	9	9	9	10	10	10
Cost per function (microcents)										
DRAM (cost/bit)	7.7	5.4	3.8	2.7	1.9	1.4	0.96	0.34	0.12	0.042
CP-MPU (cost/transistor)	107	75	53	38	27	19	13.3	4.71	1.66	0.590
HP-MPU (cost/transistor)	97	69	49	34	24	17	12	4.31	1.52	0.540
Test cost ($K/pin)										
Volume tester cost per high-frequency signal pin (HP-ASIC)—maximum	4.0	3.0	3.0	3.0	3.0	3.0	3.0	4.0	4.0	4.0
Volume tester cost per high-frequency signal pin (HP-ASIC)—minimum	1.0	1.0	1.0	1.0	1.0	1.0	1.0	2.0	3.0	4.0
Volume tester cost/pin (CP-MPU)	7.0	6.5	6.0	5.5	5.0	4.5	4.0	4.0	2.0	1.5
Power supply voltage (V)										
Vdd (high performance)	1.1	1.0	1.0	1.0	0.9	0.9	0.7	0.6	0.5	0.4
Vdd (low operating power, high Vdd transistors)	1.2	1.2	1.1	1.1	1.0	1.0	0.9	0.8	0.7	0.6
Vdd (low standby power, high Vdd transistors)	1.2	1.2	1.2	1.2	1.2	1.2	1.1	1.1	1.1	1.1
Allowable maximum power										
High-performance with heatsink (W)	130	140	150	160	170	180	190	218	251	288
Cost-performance (W)	61	75	81	85	92	98	104	120	138	158
Battery (W)—(handheld)	2.4	2.6	2.8	3.2	3.2	3.5	3.5	3.0	3.0	3.0

tions per clock cycle). The need for a progressively higher operational frequency associated with an increasing average chip size will continue to fuel the development of novel process, design, and packaging techniques.

Table 1 reflects these considerations. The highest frequency obtainable in each product generation is related to the intrinsic transistor performance (on-chip, local clock), and this relationship becomes even more direct as microarchitectural knobs (for example, pipelining) become fully exploited.

To optimize signal and power distribution across the chip, the number of interconnect layers is likely to continue to increase. As interconnect size downscaling also continues, the chip fabrication process will adopt wider use of copper (low resistivity) and various intermetal insulating materials of progressively lower dielectric constant ($\kappa \sim 2$-3). Designers will also use multiplexing techniques to increase the chip-to-board operating frequency (off-chip).

In general, signal propagation becomes more difficult due to increased capacitive and inductive coupling, which degrades edge rates and causes both timing uncertainty and potential logic errors. Additional signal degradation is associated with the inductance of wire bonds and package leads. Direct chip attachment may eventually be required for adequate mitigation of parasitic effects caused by the package.

Cost

Table 1 also shows cost trends. The ability to reduce the cost per function by 25 to 30 percent each year is a unique feature of the semiconductor industry and is the fundamental engine behind its growth. In support of this cost reduction, R&D and manufacturing require a continuously increasing financial investment. Even on a per-factory basis, the capital cost of manufacturing continues to escalate.

However, the 2001 Roadmap indicates that logic transistor size is improving only at the rate of the

lithography (0.7 times linearly and 0.5 times area reduction every technology node). Therefore, to keep the MPU chip sizes flat, the number of transistors can double only every technology node. Because the technology node rate is projected to return to a three-year cycle after 2001, the transistors per MPU chip can double only every three years after 2001. DRAM memory bit cell design improvements are also slowing down, and the bits-per-chip rate will also be slowing in the future to keep chip sizes under control. To compensate for the decrease in DRAM and MPU functions per chip, there will be increasing pressure to find alternative enhancements from the equivalent productivity scaling benefits of chip and system-level architecture and designs.

Even though the rate of increase in on-chip complexity could slow in the future, the number of functions per chip will continue to grow. Increased chip functions drive an increase of test-method complexity, which in the past resulted in nonlinear cost increases to manufacturing test in capital for additional ATE (automated test equipment) hardware and longer device test times. Even though ATE cost-per-pin is forecast to decline, this is more than offset by increased device pin counts and complexities. Built-in self-test (BIST) and design for testability (DFT) must move forward to enable critical manufacturing test cost scaling—for example, reduced-pin-count ATEs.

Meeting these challenges will require advances on all fronts—particularly new front-end processes that overcome the limitations of current complementary metal-oxide semiconductor (CMOS) technology—but here we focus on challenges in the design and test arena that is more relevant to *Computer*'s readership.

SYSTEM DRIVERS FOR DESIGN

Previous ITRS editions focused on MPUs, DRAM, and application-specific integrated circuit (ASIC) product classes, with only cursory mention of system-on-chip (SoC) and analog/mixed-signal (AMS) circuits. The unstated assumption was that technological advances only needed to be linear and that all semiconductor products would deploy them. Today, the introduction of new technology solutions is increasingly application driven, with products for different markets using different combinations of technologies at different times: Battery-powered mobile devices are replacing wall-plugged servers, and SoC and system-in-package designs that incorporate building blocks from multiple sources are supplanting in-house, single-source chip designs.

The 2001 ITRS updates and more clearly defines the set of system drivers that previous ITRS editions used, providing quantitative, internally self-consistent models that support extrapolation and adapt more smoothly to future technology developments. Due to DRAM's well-understood commodity nature, the ITRS focuses on high-volume custom microprocessors, AMS, and SoC drivers.

As the number of functions per chip increases, testing the final products becomes increasingly difficult and more costly.

Custom MPUs

High-volume custom MPUs incorporate the most aggressive design styles and manufacturing technologies. It is for these high-volume parts that developers make changes to the manufacturing flow, create new design styles and supporting tools (the large revenue streams can pay for new tool creation), and uncover subtle circuits issues. Thus, while developing custom MPU designs is extremely labor intensive, they offer new design and fabrication technology and new automation methods that the entire industry leverages.

MPUs are part of the segment that drives integration density and design complexity, the power-speed performance envelope, large-team design process efficiency, test and verification, power management, and packaged system cost. Historically, there have been two types of MPUs over the course of the Roadmap: cost-performance (CP) desktop and high-performance (HP) servers, with constant die areas of 140 mm^2 and 310 mm^2, respectively.

In contrast to previous ITRS models, the core message in the 2001 model is that power and cost are strong limiters of die size. Future MPUs will likely require a merged desktop-server category (the distinction is already blurred today) and a mobile category (essentially a low-power, high-performance SoC).

Design productivity, power management, multicore organization, I/O bandwidth, and circuit and process technology are key contexts for the future evolution of the traditional MPU.

Design productivity. The complexity and cost of design and verification of MPU products have rapidly increased to the point where developers devote thousands of engineer-years (and a design team of hundreds) to a single design, yet processors reach market with hundreds of bugs.

Power management. Power dissipation limits of cost-effective packaging, estimated to reach 50 W per cm^2 for forced-air cooling by the end of the Roadmap, cannot continue to support high-supply voltages. Historically, these voltages scale at 0.85 times per generation instead of 0.7 times ideal scaling, and frequencies historically scale by 2 times per

Power dissipation limits of cost-effective packaging cannot continue to support high-supply voltages.

generation instead of the ideal 1.4 times.

Past MPU system driver clock frequency trends were interpreted as future CMOS device performance (switching speed) requirements, leading to large off-currents and extremely thin gate oxides. Given such devices, MPUs that simply continue using existing circuit and architecture techniques would exceed package power limits by a factor of more than 25 times by the end of the Roadmap. Alternatively, MPU logic content or logic activity would need to decrease to match package constraints. Portable low-power embedded systems have more stringent power limits and will encounter such obstacles even earlier than MPUs.

Power efficiencies are up to four orders of magnitude greater for direct-mapped hardware than for general-purpose MPUs, and this gap is increasing. As a result, traditional processing cores will face competition from application-specific or reconfigurable processing engines for space on future SoC-like MPUs.

Multicore organization. In an MPU with multiple cores per die, the cores can be smaller and faster to counter global interconnect scaling, and developers can optimize them for reuse across multiple applications and configurations.

In addition to allowing power savings, multicore architectures may exploit redundancy to improve manufacturing yield. Future MPU organization will likely increase the on-chip memory hierarchy, which, if only in a relatively trivial way, affords better control of leakage and total chip power. Evolutionary microarchitecture changes—superpipelining, superscalar, predictive methods—appear to be running out of steam. Thus, more multithreading support will emerge for parallel processing, as well as more complex hardwired functions or specialized engines for networking, graphics, security, and so forth. Flexibility-efficiency tradeoffs shift away from general-purpose processing.

I/O bandwidth. In MPU systems, I/O pins mainly connect to both high-level cache memory and main-system memory. Increased processor performance has been pushing I/O bandwidth requirements. L2 or L3 caches traditionally use the highest-bandwidth port, but recent designs integrate the memory controller on the processor die to reduce memory latency. These direct memory interfaces require more I/O bandwidth than the cache interface.

Many designs replace the system bus with high-speed point-to-point interfaces that require much faster I/O design, exceeding gigabit-per-second rates. While serial links have achieved these rates for a while, integrating a large number of these I/Os on a single chip presents challenges for design (each circuit must be very low power), test (the tester needs to run this fast), and packaging (packages must act as balanced transmission lines, including the connection to the chip and the board).

Circuit and process technology. The growing process variability implicit in feature size and device architecture roadmaps, including thinner and less reliable gate oxides, subwavelength optical lithography requiring aggressive reticle enhancement, and increased vulnerability to atomic-scale process variability, severely threatens parametric yield (dollar per wafer after bin-sorting). This will require more intervention at the circuit and architecture design levels.

While using dynamic circuits is attractive for performance in lower-frequency or clock-gated regimes, noise margin and power dissipation concerns may limit this approach. Error-correction for single-event upset in logic will increase, as will using redundancy and reconfigurability to compensate for yield loss. Power management will require using a combination of techniques from several component technologies. Application, OS, and architecture optimizations include parallelism, adaptive voltage, and frequency scaling. The increased use of silicon-on-insulator techniques is a process innovation. Circuit design optimization techniques include the simultaneous use of multi-Vth, multi-Vdd, minimum-energy sizing under throughput constraints, and multidomain clock gating and scheduling.

Analog and mixed-signal designs

AMS designs include RF, analog, and analog-to-digital and digital-to-analog converters. At least part of the AMS chip needs to measure signals with high precision. Because analog chips have very different design and process technology demands than digital circuits, scaling them into new technologies is a difficult challenge. While technology scaling is always desirable for digital circuits due to reduced power, area, and delay, it is not necessarily helpful for analog circuits in which dealing with precision requirements or signals from a fixed voltage range is more difficult. In general, AMS circuits (for example, RF and embedded passives) and process technologies (for example, silicon-germanium) present severe challenges to cost-effective CMOS integration.

The need for precision also affects tool requirements for analog design. Digital circuit design follows a set of rules that allow logic gates to function correctly: As long as the design follows these rules, precise calculation of exact signal values is not

needed. Analog designers, on the other hand, must be concerned with a number of second-order effects to obtain the required precision. Relevant issues include coupling (capacitance, inductance, and substrate) and asymmetries (local variation of supplies, as well as implantation, alignment, etching, and other fabrication effects). Analysis tools for these issues are mostly in place but require expert users; synthesis tools are at best preliminary. Manufacturing test for AMS circuits is essentially unsolved.

For most of today's mixed-signal designs—particularly classical analog designs—a voltage difference represents the processed signal, and the supply voltage determines the maximum signal. The most daunting mixed-signal challenges are

- *decreasing supply voltage*, which requires current-mode circuits, charge pumps for voltage enhancement, and thorough optimization of voltage levels in standard-cell circuits;
- *increasing relative parametric variations*, which requires active mismatch compensation and tradeoffs of speed versus resolution;
- *increasing numbers of analog transistors per chip*, which requires faster processing speed and improved convergence of mixed-signal simulation tools;
- *increasing processing speed (clock frequencies)*, which requires more accurate modeling of devices and interconnects, as well as test capability and package- and system-level integration;
- *increasing leakage and crosstalk arising from SoC integration*, which requires more accurate crosstalk and delay modeling and fully differential design for RF circuits; and
- *shortage of design skills and productivity* arising from lack of training and poor automation, which requires education and basic design tools research.

An ideal design process would reuse existing mixed-signal designs and adjust parameters to meet interface specifications between a given SoC and the outside world. However, such reuse depends on a second type of MOSFET (metal oxide semiconductor field-effect transistor) that does not scale its maximum operating voltage. This has led to the Roadmap's specification of a mixed-signal CMOS transistor that uses a higher analog supply voltage and stays unchanged across multiple digital technology generations. Even with such a device, however, voltage reduction and development time of analog circuit blocks are major obstacles to low-cost and efficient scaling of mixed-signal functions.

The Roadmap specifies a mixed-signal CMOS transistor that uses a higher analog supply voltage and stays unchanged across multiple digital technology generations.

System-on-chip design

A yet-evolving product class, SoC design integrates pieces of technology from other system driver classes—for example, MPU, memory, AMS, and reprogrammable fabrics—into a wide range of high-complexity, high-value semiconductor products. Typically, SoC manufacturing and design technologies were originally developed for high-volume custom drivers. Since reduced design costs and higher levels of system integration are its principal goals, the SoC driver class most closely resembles the ASIC category.

The primary difference between ASIC and SoC designs is that SoCs emphasize reusing intellectual property (IP) to improve productivity. In addition, SoC integration potentially encompasses heterogeneous technologies. SoCs reuse both analog and high-volume custom cores as well as blocks of software technology. The primary benefit of SoC designs is that reusing blocks is more efficient and cost effective than using equivalent from-scratch designs.

Cost considerations drive the deployment of low-power process and low-cost packaging solutions, along with fast turnaround time (TAT) design methodologies. The latter, in turn, require new standards and methodologies for IP description, IP test (including BIST and self-repair), block interface synthesis, and so forth. In addition to the need for chip-package cooptimization, integration considerations drive the demand for heterogeneous technologies such as flash, DRAM, MEMS, ferroelectric RAM (FERAM), magnetoresistive RAM (MRAM), and chemical sensors that implement particular system components. Thus, SoC is the driver for convergence of multiple technologies not only in the same system package, but also potentially in the same manufacturing process.

Because SoC designs offer low-cost, rapid system implementation, power management and design productivity have important implications for the achievable design space. The 2001 ITRS defines a prototypical low-power SoC (LP-SoC) PDA application and applies two analyses to obtain future power management requirements. The first analysis accepts the system specifications (0.1 W peak power and 2.1 mW standby power) in a top-down fashion. The second approach derives the power requirements bottom-up from the implied logic and memory content, as well as process and circuit para-

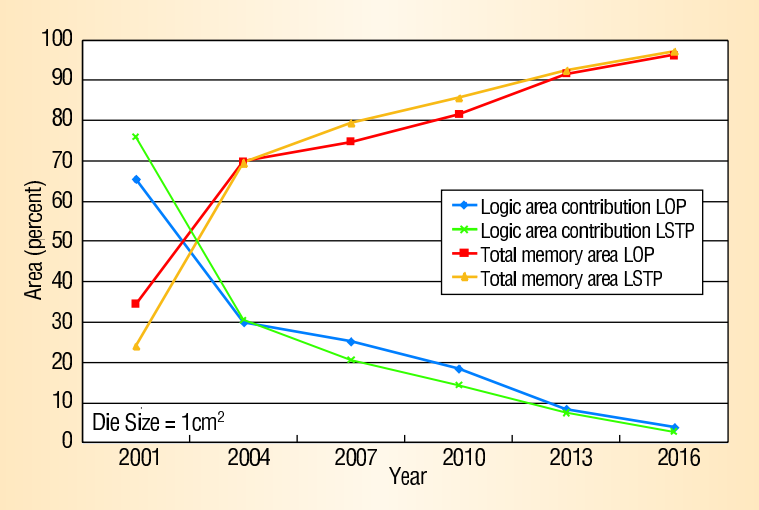


Figure 2. Power gap effect on chip composition. Memory content outstrips logic content faster with LSTP devices because they have much higher operating power than LOP devices.

meters. Table 1 shows power constraints projected through 2016.

Figure 2 projects logic/memory composition of LP-SoC designs, assuming that chip power is constrained according to a power budget of 0.1 W and that chip size is constrained to 100 mm^2. Memory content outstrips logic content faster with LSTP (low standby power) devices because they have much higher operating power than LOP (low operating power) devices. Without substantial improvements in power management capability, memory will asymptotically dominate both models by 2016. Given the projection that PDA chip size will grow at approximately 20 percent per node even though power remains flat at 0.1 W, this would lead to even more extreme memory-logic imbalances in the long term.

DESIGN

The overriding message in the 2001 Roadmap is that design cost is the greatest threat to continuation of the semiconductor industry's phenomenal growth. Manufacturing nonrecurring engineering (NRE) costs are just reaching $1 million (mask set and probe card), whereas design NRE costs routinely reach tens of millions of dollars. We measure manufacturing cycle times in weeks, with low uncertainty, whereas we measure design and verification cycle times in months or years, with high uncertainty. Moreover, design shortfalls are responsible for silicon respins that multiply manufacturing NRE costs.

Despite an acknowledged design productivity gap in which the number of available transistors grows faster than the ability to design them meaningfully, investment in process technology has by far dominated investment in design technology. The good news is that developers continue to make progress in design technology (DT): The estimated design cost of a low-power SoC PDA was approximately $15 million in 2001 versus $342 million if DT innovations had not occurred between 1993 and 2001. The bad news is that software now routinely accounts for 80 percent of embedded-systems development cost; test cost has grown significantly relative to total manufacturing cost; verification engineers are twice as numerous as design engineers on microprocessor project teams—and the list goes on. In 2001, many previous design technology gaps became crises.

Complexity challenges

DT faces two basic types of complexity: silicon and system. Silicon complexity refers to the impact of process scaling and the introduction of new materials or device/interconnect architectures. Previously ignorable phenomena (implied challenges) now have greater impact on design correctness and value, including:

- *nonideal scaling of device parasitics and supply/threshold voltages*—leakage, power management, circuit/device innovation, current delivery;
- *coupled high-frequency devices and interconnects*—noise/interference, signal integrity analysis and management;
- *manufacturing equipment limits*—statistical process modeling, library characterization;
- *scaling of global interconnect performance relative to device performance*—communication, synchronization;
- *decreased reliability*—gate insulator tunneling and breakdown integrity, joule heating and electromigration, single-event upset, general fault tolerance;
- *complexity of manufacturing handoff*—reticle enhancement and mask writing/inspection flow, NRE cost; and
- *process variability*—library characterization, analog and digital circuit performance, error-tolerant design, layout reuse, reliable and predictable implementation platforms.

Silicon complexity places long-standing paradigms at risk: System-wide synchronization becomes infeasible due to power limits and the cost of robustness under manufacturing variability; the CMOS transistor becomes subject to ever-larger statistical variabilities in its behavior; and fabrication of chips with 100 percent working transistors and interconnects becomes prohibitively expensive.

System complexity refers to exponentially increasing transistor counts enabled by smaller feature sizes and spurred by consumer demand for increased functionality, lower cost, and shorter time to market. Implied challenges include:

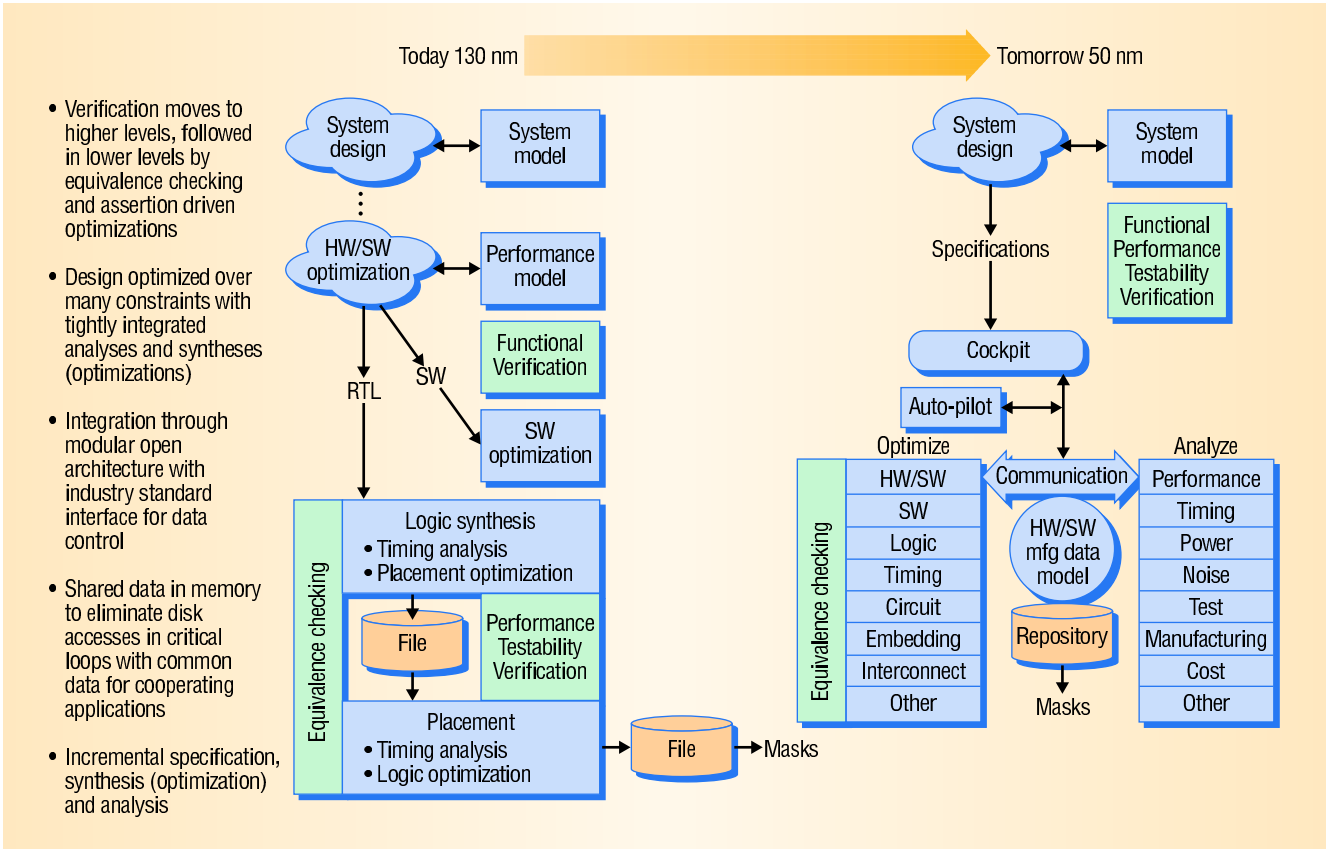


Figure 3. Evolution of design system architecture into an integrated system wherein logical, physical, layout, and other tools can operate together.

- *reuse*—support for hierarchical design, heterogeneous SoC integration-modeling, simulation, verification, and component block test—especially AMS;
- *verification and test*—specification capture, design for verifiability, verification reuse for heterogeneous SoCs, system-level and software verification, AMS and novel device verification, test access, self-test, intelligent noise/delay fault testing, tester timing limits, test reuse;
- *cost-driven design optimization*—manufacturing cost modeling and analysis, quality metrics, cooptimization at die-package-system levels, optimization with respect to multiple system objectives such as fault tolerance and testability;
- *embedded software design*—predictable platform-based system design methodologies, codesign with hardware and for networked system environments, software verification/analysis;
- *reliable implementation platforms*—predictable chip implementation onto multiple circuit fabrics, higher-level handoff to implementation; and
- *design process management*—design team size and geographic distribution, data management, collaborative design support, "design through system," supply chain management, metrics, and continuous process improvement.

Together, silicon and system complexity trends lead to superexponentially increasing design process complexity. To combat this complexity, eight overarching methodology precepts are called out for the future evolution of DT: exploit reuse; evolve DT rapidly; avoid iterations; replace verification by prevention; improve predictability; orthogonalize concerns (for example, by separating behavior from architecture, or computation from communication); expand the scope of DT (up to package and board levels, down to mask and process, from digital hardware to software and AMS, and so on); and unify previously disparate subareas of DT. Figure 3 shows the transition of design system architecture in light of these precepts: the traditional waterfall, in which design proceeds independently at discrete levels, evolves into an integrated system wherein logical, physical, layout, and other tools can operate together.

Crosscutting challenges

The Roadmap sets out detailed challenges with respect to five traditional areas of DT: design process; system-level design; logic, circuit, and physical design; design verification; and test. However, beyond enumerating these detailed challenges, the 2001 Roadmap also identifies five crosscutting challenges that encompass all relationships between electronic design automation and the other industries that support the semiconductor industry whose solutions are distributed across all

A basic DT challenge is to improve characterization, modeling, and analysis and estimation of noise and interference at all levels of design.

areas of design technology. Our strong hope is that challenges impossible to solve within a single ITRS technology area are solvable with design technology partnership. For example, more rapid adoption of new fault models for crosstalk and path delay, along with corresponding automatic test pattern generation (ATPG) and BIST techniques, might address test equipment and speed limitations.

The five crosscutting challenges are productivity, power, manufacturing integration, interference, and error tolerance.

Productivity. To avoid exponentially increasing design costs, overall productivity of designed functions on chip—as well as reuse productivity (including migration) of design, verification, and test—must scale at more than two times per node. Verification has become a bottleneck that has reached crisis proportions, calling for reliable and predictable silicon implementation fabrics that support higher-level system design handoffs and, particularly in the SoC arena, automated methods for AMS synthesis, verification, and test. Reducing DT time to market requires standards that promote stability, predictability, and interoperability.

Power. Nonideal scaling of planar CMOS devices, together with the Roadmap for interconnect materials and package technologies, presents a variety of power management and current delivery challenges. MPU power dissipation will exceed high-performance single-chip package power limits by 25 times at the end of the Roadmap, whereas LP-SoC PDA drivers require flat average and standby power even as logic content and throughput continue to grow exponentially. DT must address the resulting power management gap in which increasing power densities worsen thermal impact on reliability and performance and decreasing supply voltages worsen switching currents and noise. These trends stress on-chip interconnect resources, test equipment power delivery and dynamic response limits, and even current latent defect acceleration paradigms.

Manufacturing integration. Feasibility of future technology nodes will depend on sharing challenges within the industry as a whole. Die-package-board cooptimization and analysis may improve system implementation cost, performance verification, and overall design TAT as well as system-in-package DT. New DT for correctness under manufacturing variability—for example, variability-aware circuit design, design for regularity, timing-structure optimization, and static-performance verification—may relax critical-dimension control requirements in the lithography, process integration, devices, and structures, front-end processing, and interconnect technology areas. Finally, more intelligent interfaces that mask production and inspection flows may reduce manufacturing NRE costs.

Interference. Noise and interference increasingly hamper resource-efficient communication and synchronization, which global interconnect scaling trends already challenge. Prevailing signal integrity methodologies in logical, circuit, and physical design—while apparently scalable through the 100 nm node—are reaching their limits of practicality. These methodologies include repeater insertion rules for long interconnects, slew-rate control rules, and power/ground distribution design for inductance management.

Scaling and SoC integration of mixed-signal and RF components will require more flexible and powerful methodologies. Issues include noise headroom (especially in low-power devices and dynamic circuits); large numbers of capacitively and inductively coupled interconnects; supply voltage IR drop and ground bounce; thermal impact on device off-currents and interconnect resistivities; and substrate coupling. A basic DT challenge is to improve characterization, modeling, and analysis and estimation of noise and interference at all levels of design.

Error tolerance. Error tolerance, correction, and self-repair could dramatically increase manufacturing yields but will require additional effort in verification and test. Technology scaling likely forces such a paradigm shift, which leads to more transient and permanent failures of signals, logic values, devices, and interconnects. Below 100 nm, single-event upsets (soft errors) severely impact both memory and logic field-level product reliability. Atomic-scale effects demand new "soft" defect criteria, such as for noncatastrophic gate oxide breakdown.

In general, automatic insertion of robustness into the design will become a priority as systems become too large to functionally test at manufacturing exit. Potential measures include automatic introduction of redundant logic and on-chip reconfigurability for fault tolerance, development of adaptive and self-correcting or self-repairing circuits, and software-based fault tolerance.

TEST

For many years, at-speed functional test has provided a robust methodology for high-volume man-

ufacturing to achieve the required outgoing quality levels. However, it now appears that this method is running out of gas for several reasons, not the least of which are geometrically increasing test-development engineering resources and increasing ATE cost. Manufacturing yield loss associated with the at-speed functional test methodology is related to the growing gap between ATE performance and ever-increasing device I/O speeds that require increased accuracy for proper resolution of timing signals.

While semiconductor off-chip speeds have improved at 30 percent per year, tester accuracy has improved at a rate of only 12 percent per year. Typical headroom offered by testers five times faster than device speeds in the 1980s have disappeared. If the current trends continue, tester-timing errors will approach the cycle time of the fastest devices. In 2001, yield losses due to tester inaccuracy were already becoming a problem when using a traditional functional test methodology during manufacturing.

Moreover, even if upgrading or replacing manufacturing test equipment with each increase in device performance were affordable, avoiding manual test writing in the functional test environment—which requires tens of person-years for highly complex designs—has proven to be impossible. As a consequence, the search for low-cost design-for-test (DFT) equipment solutions has recently generated significant industry momentum. DFT techniques like scan and BIST can enable automatic test-content generation and, at the very least, drastically reduce the manual test-writing task irrespective of the potential capital cost savings. Highly integrated SoC devices require a highly structured DFT approach to enable reuse of test collateral and avoid geometric or exponential growth of the test development and validation effort and test manufacturing cost.

Design for test

The 1999 Roadmap provided the first focused requirements definition for low-cost DFT testers, and the 2000 update further clarified the requirements of this paradigm shift for device DFT and manufacturing test equipment. The extensive collaboration between semiconductor manufacturers and test equipment suppliers during the process of generating these requirements has demonstrated that highly custom individual designs use DFT methodologies that converge toward a common set of tester building blocks. This significant conclusion builds confidence that ATE developers can design and configure generic DFT equipment that meets the industry's requirements, mitigating the need for custom solutions. DFT-based approaches require continued research to increase coverage of process defects by developing advanced methodologies to apply patterns based on existing fault models to designs and identifying novel fault models.

New I/O protocols operate in the multigigahertz range and are significantly more complex.

Even so, nanometer process technology, increasing clock rate, and SoC integration present severe challenges and may limit the application of conventional DFT techniques. SoC designs are breaking the traditional barriers between digital, memory, analog, RF, and mixed-signal test equipment requirements, resulting in a trend toward highly configurable, one-platform-fits-all test solutions. Increasing demand for bandwidth at the system level and constant or shrinking final package form-factor are driving wide proliferation of new high-speed serial protocols for off-chip communication across device types. The analog nature of these interfaces and the demand for device interoperability drive extensive at-speed parametric test requirements and new test and DFT methods into manufacturing.

While DFT methodologies are feasible in these areas, it is expected that this technology will continue to lag behind leading-edge device performance and complexity.

Highly integrated SoC designs. Integration of preexisting design blocks into larger integrated devices produces nonlinear complexity growth for design tools, DFT, and manufacturing test—even when the blocks are homogeneous (for example, all logic). Increasingly, devices combine analog, mixed-signal, and nonvolatile flash with logic and RAM. Traditional test methods and ATE equiment are radically different across these device types whereas silicon complexity and costs are relatively predictable for integrated devices. Therefore, embedded blocks and mixed-device types drive highly nonlinear and unpredictable increases in testability, design verification, and manufacturing test costs.

ASIC or MPU macros wholly embedded within larger logic devices are seeing this impact, with manufacturing test costs in some cases already exceeding silicon costs. Even with DFT, these costs may be nonlinear. Direct-access DFT (DAT) testing of embedded blocks may also entail an order of magnitude longer test time than testing non-embedded versions, ultimately driving a much wider adoption of BIST than previously seen.

Test will continue to leverage functional methodology to obtain the coverage required to guarantee outgoing product quality.

Larger portions of test will require expanded DFT techniques and protocols—for example, IEEE P1500—as well as significant use of BIST or embedded software-based self-testing to counteract the growth in test complexity due to increasing design integration levels.

High-speed device interfaces. Component I/O speed has become as important to system performance as core clock frequency or transistor and architectural performance. New I/O protocols operate in the multigigahertz range and are significantly more complex—with source-synchronous, differential, and even simultaneous bidirectional schemes operating at gigabit-per-second rates and with differential voltage swings a fraction of the supply Vdd range.

ATE and component-test legacies include common clock-based testing and I/O measurements in the megahertz range. Hence, I/O speeds and protocols drive significant instrumentation, materials, and cost challenges for the ATE equipment, interface hardware, and test sockets needed by both design verification and manufacturing test. This inflection point demands broad industry development and application of on-die testability capabilities specifically for I/Os. I/O DFT and BIST methods such as loopback, jitter measurement, and edge detection will become standard techniques for verification and manufacturing test of these new I/O architectures. Without DFT innovation, it is conceivable that the trend toward protocol-based high-speed I/O could drive ATE toward protocol-specific test solutions—a complex and expensive proposition.

Reliability screens. Manufacturing test has historically not only measured device performance and functionality, but also performed the required business task of identifying and segregating latent reliability defects or, more specifically, defect-driven reliability failures. The exponentially increasing (approximately 10 times per technology node) leakage currents of advanced silicon technologies severely limit dynamic burn-in, IDDQ (direct drain quiescent current), and above-Vdd-voltage stress during test.

The decreasing ratio of stress voltage to nominal Vdd limits the acceleration, identification, and screening capabilities of both burn-in and on-ATE-voltage-stressing. At 180 nm and 130 nm, thermal runaway limits the use of temperature acceleration methodologies and drives nonlinear cost increases for burn-in for high-end products such as microprocessors. At the same time, with existing and forecast trends for increasing device leakage and background currents, advanced IDDQ techniques such as IDDQ delta are becoming extremely limited due to the difficulty in identifying the IDDQ "signal" within the background current noise.

In the near term, significant manufacturing cost increases could result from the yield impact (overkill) and equipment cost of extending current approaches just to keep pace with market reliability requirements.

Whither functional test?

Does directing the industry momentum toward DFT-based designs to decrease product test cost make functional test go away? As technology has evolved over time, functional test equipment costs have decreased with respect to a fixed capability and have held roughly constant for leading-edge performance. Test will continue to leverage functional test methodology as one opportunity to obtain the coverage required to guarantee outgoing product quality. However, it is expected that DFT will be used when needed to limit the functional test performance envelope in production by reducing I/O data rate requirements, enabling low pin-count testing and reducing the dependence on expensive instruments. DFT will let manufacturers step off the technology treadmill associated with functional test equipment and enable greater reuse of this equipment for manufacturing test across technology nodes.

In the device debug and characterization world, at-speed functional and analog test will continue to serve as a primary vehicle for detecting the root cause of design and process errors and marginalities. At the same time, traditional test equipment-based methodologies will need to correlate DFT-based results to end-use environment conditions. However, it is not expected that this equipment will proliferate into manufacturing, but rather that it will be used to prove manufacturing capability on lower cost high-volume testers.

This represents a significant challenge to the industry: Should this trend continue, it would reduce the total available market for the most complex, development-intensive test equipment. It is unclear whether or not there is a compelling business model to develop this equipment without a dramatic increase in capital cost. Thus, avoiding rising equipment costs requires identifying new methodologies for design debug and characterization.

Failure analysis

One of the most critical challenges for the ITRS Roadmap is the continuing success of device and defect failure analysis (FA) to drive the rapid yield learning that enables succeeding technology generations. The migration of CMOS technology towards 65-nm feature sizes will severely challenge the traditional FA process as device features and defects—already below the optical wavelength for microscopic techniques—become even more reclusive to even advanced physical FA technologies.

Physical FA equipment costs and throughput are increasing dramatically, threatening the expected yield learning rate the Roadmap requires to continue to move forward. Hardware-based physical FA—that is, deprocessing and physical characterization/inspection—will remain essential, but will become critically dependent on software and on-die diagnostic techniques for fault localization. Increased circuit sensitivity requires finding smaller, more subtle defects. Tighter pitches and smaller device features require greater spatial resolution. Flip-chip packaging and increasing numbers of metal layers force the increased use of backside analysis techniques, thereby also increasing the throughput to actual FA results.

These factors will combine to make the physical FA process too slow and difficult to be relied upon as a routine analysis procedure. The need for software-based fault localization, a key alternative/supplement to traditional hardware-based fault localization, is especially acute, requiring significant breakthroughs.

These are but a few of the difficult and exciting challenges facing the design and test communities and the semiconductor industry as a whole. To continue its phenomenal historical growth and continue to follow Moore's law, the semiconductor industry will require advances on all fronts—from front-end processes and lithography to assembly and packaging and factory integration and, increasingly, to design and test innovation. The Roadmap's goal is to bring together experts in each of these fields to determine what those challenges are and potentially how to solve them.

Although we make no pretense of having fully characterized these challenges here, our intent is to give readers a hint of the vast collaborative effort behind these potentially novel solutions and to refer them to the ITRS Web site for details of the full report (http://public.itrs.net/). ■

Acknowledgments

We acknowledge the efforts of the many individuals who contributed to making the 2001 edition of The International Technology Roadmap for Semiconductors a successful endeavor.

Alan Allan *is a staff engineer at Intel. His research interests are industry technology and economic roadmapping and modeling. He received a BSEE from the University of Colorado at Boulder. Contact him at alan.k.allan@intel.com.*

Don Edenfeld *is responsible for test equipment definition and selection at Intel. He received a master's degree in electrical engineering from the University of Virgina. He is a member of the IEEE Computer Society and is the 2001-2002 ITRS Test TWG chair. Contact him at donald.e.edenfeld@intel.com.*

Andrew B. Kahng *is a professor of computer science engineering and electrical and computer engineering at the University of California, San Diego. He received a PhD in computer science from the University of California, San Diego. He is a member of the ACM, the IEEE, and SPIE. Contact him at abk@ucsd.edu.*

William H. Joyner Jr. *is a research staff member at IBM on assignment as director of CAD and test at the Semiconductor Research Corporation. He is cochair of the US Design TWG for the 2001 ITRS. Joyner received a PhD in applied mathematics from Harvard University. He is a Fellow of the IEEE. Contact him at joyner@src.org.*

Mike Rodgers *is responsible for test technology integration at Intel. He received at BS in electrical engineering from the University of Illinois. He is cochair of the ITRS Test TWG. He is a member of the IEEE Computer Society. Contact him at michael.j.rodgers@intel.com.*

Yervant Zorian *is the vice president and chief scientist of Virage Logic and chief technology advisor for LogicVision in San Jose, Calif. He received a PhD in electrical engineering from McGill University. Zorian, a Fellow of the IEEE and a member of the IEEE Computer Society, is the editor in chief emeritus of* IEEE Design and Test. *Contact him at zorian@logicvision.com.*